An Animal Model of the Childhood Lymphoma Implicates Important Prognostic Genes in Cancer Survival

Rieko Sotojima

Submitted to the
College of Nursing
in Partial Fulfillment of the Requirements for the Degree of

Bachelor of Science in Nursing with Research Distinction

At

The Ohio State University

May 4, 2018

Thesis Advisor:
Jennie Rowell PhD, RN
Assistant Professor
College of Nursing Center of Excellence in Critical and Complex Care
390 Newton Hall, 1585 Neil Ave., Columbus, OH 43210
Office Phone: (614) 292-4497
Email: Rowell.14@osu.edu

Second Reader:
Shannon Gillespie, PhD, RN
Assistant Professor
358 Newton Hall, 1585 Neil Avenue Columbus, OH 43210
Center for Women, Children, and Youth, College of Nursing
The Ohio State University
Office Phone (614) 292-4589
Email: Gillespie.175@osu.edu

ABSTRACT

Background: In the US, cancer is 2nd leading cause of death in children. Surprisingly, the prevailing treatment regimens have been modified from adult therapies. Yet, substantial differences exist between adult and pediatric cancers. Determining the underlying genetic causes of pediatric cancers is problematic for many reasons, including the overall rarity. In order to gain tractability in this disease, we need a new model that closely recapitulates pediatric cancers, but at much higher incidence. The pet dog is ideally suited to be this model.

Significance: Children experience varying degrees of suffering due to the side effects of treatment modalities that are primarily non-specific. Isolating genetic causes would allow for personalized treatments and increase survival.

Purpose: The purpose of this work was to identify genetic and epigenetic (DNA methylation) biomarkers in an animal model of pediatric lymphoma that are associated cancer survival.

Conceptual Framework: We used the genetic single gene-two hit model (Hussian, 2015) to guide this work. This framework combines rare and common variant theories with the addition of mutator/anti-mutator modulation to lead to disease development.

Methods: Using chemotherapy naive tumor samples from 71 dogs diagnosed with aggressive lymphomas, we analyzed gene expression using Affymetrix Genechip 2.0 arrays. Additionally, we performed whole genome methylation analysis (MeDIP) on a subset of 6 dogs. Data was normalized and analyzed using a genomics statistical suite (JMP). We further validated our expression results in an online publically available database of 91 pediatric lymphoma tumors.

Results: After selecting the significant transcripts (fold-change ≥ 3 and $p \leq 0.001$), we correlated these in a subset of dogs with survival time. We identified 6 prognostic genes significantly associated with cancer survival. In the methylation analysis, we identified 633 differentially methylated first exons of genes ($p \leq 0.01$). Overlapping both the expression and methylation analysis, we identified 1 gene (*Cadherin 1, CDH1*). The expression of this gene was significantly associated with survival in pediatric lymphoma ($p = 0.013$).

Conclusions: This study is one of the first to provide genetic understanding of aggressive lymphomas that have relevance in children. Focusing on the molecular properties of pediatric cancer will provide significant clinical utility.

AN ANIMAL MODEL OF THE CHILDHOOD LYMPHOMA IMPLICATES IMPORTANT PROGNOSTIC GENES IN CANCER SURVIVAL

INTRODUCTION

In the United States, the leading cause of death by disease after infancy among children and adolescents is cancer, with nearly 2,000 dying from the disease in 2017 alone [1]. Yet, survival has never been higher, with 89% (~13,000) surviving $\geq$ 5 years. Due to the growing nature of this population, selecting effective, risk-adapted medical treatment can be very challenging [2]. Currently, the majority of existing treatment regimens for pediatric cancers have been modified from adult therapies [3]. Within the last decade, the long-term effects of such non-specific treatments have been identified [4]. Termed "late effects", long-term childhood cancer survivors experience debilitating physical and psychosocial diseases and symptoms after completion of treatment that persists into and throughout adulthood [5-8]. In fact, nearly 90% of childhood cancer survivors experience "a lot or a great deal of suffering from at least one symptom" [9, 10]. Despite this, the burden of symptom experience during cancer diagnosis, treatment, and follow-up is not well managed, notwithstanding considerable research on the long-lasting nature of such symptoms and clinical recommendations based on this evidence [11]. Currently, implementation of these guidelines and research to understand the complex interaction of systems (physiological, psychological, environmental, etc.) that lead to the symptom experience associated with childhood cancers is not well elucidated [9, 12, 13]. Cancer is now recognized as a disease of abnormal epigenetics, and aberrant DNA methylation patterns have already been associated with some childhood cancers [14-16]. Urgently needed are clear genetic/epigenetic treatment targets, which are cancer cell-specific and do not have an effect on the rapidly growing tissues in a child. Here, we use the pet dog as a model of childhood high-grade (aggressive) lymphomas. We identify the

epigenetic biomarkers in canine lymphoma, associated with survival and cancer development, to gain higher understanding of pediatric cancers. The unique aspects of pediatric Diffuse Large B-Cell Lymphoma (DLBCL) can be used for prevention, diagnosis, treatment, and improvement of long-term outcomes for children with DLBCL. Moreover, this genetic knowledge could allow for sub-classification of tumors, symptoms, and treatments which may be utilized for translational studies in the future.

BACKGROUND

Pediatric High-grade Lymphomas

In children, Lymphoma (Hodgkin lymphoma (HL) and Non-Hodgkin lymphoma (NHL)) is the third most common childhood malignancy [18]. NHL is a heterogeneous classification of malignancies and includes more than 70 subtypes [19-21]. NHL arises from either B or T white blood cells (lymphocytes) of the immune system at various stages of cell differentiation [22, 23]. Overall, NHL accounts for about 5% of all childhood cancers, with approximately 800 new pediatric cases diagnosed each year [24-26]. Of those cases, 10-20% will be DLBCL, the most common subtype of NHL [27]. Interestingly, various age-related differences in tumor biology and survival have been observed suggesting that childhood cancers differ significantly in critical ways from adult cancers (Table 1)[28].

Diffuse Large B-Cell Lymphoma			
Category	**Adult**	**Pediatric**	**Canine**
Frequency of New NHL/Year	> 21,000	7,800 - 14,000	100,000 - 600,000 (estimated)
C-Myc Expression	Decreased	Increased	Increased
BCL-2 Expression	Increased	Decreased	Decreased
ABC Phenotype	30%	15-40%	50%
GC Phenotype	48%	60-85%	50%
Morphology	Multicentric	Centroblastic	Centroblastic

Table 1: Comparison of DLBCL in Adults, Children, and Dogs. We compared adults and children on several cancer related features and find adults differ significantly from kids but kids and dogs share features.

For example, one significant difference between pediatric and adult DLBCL is the prognosis; while adults who receive multi-agent chemotherapy only has a survival rate of 50%, children have the survival rate of 85-95% [27]. Further, differential expression in pediatric and adult DLBCL suggest mechanistic reasons for the disparate outcomes [22]. *Bcl2* (apoptosis regulator gene) is highly expressed in adult DLBCL while children with the disease have a low expression [27]. Conversely, pediatric cases demonstrate increased expression of *c-Myc* (a proto-oncogene) while adults cases demonstrate low expression [27]. Additionally, extranodal disease is more frequent in pediatric cancers, the tumor morphology differs between that of adults and children, and in pediatric DLBCL a common translocation in the adult cancer [t(14;18)] is rarely observed in the pediatric cases [22, 27].

Yet, despite the clear differences in the nature of pediatric and adult onset DLBCL, there are many aspects of pediatric DLBCL genetics that remain unclear due to the inherent issues associated with studying pediatric cancers. For example, while cancer is the 2nd leading cause of death in children, incidence of cancer in kids is rare overall, especially given the multiple types of cancer that children develop [29]. The most common subtype of NHL only accounts for a total of 80-160 new diagnoses each year [27]. Thus, dividing pediatric cancers into specific subtypes for meaningful clinical intervention results in insufficient sample sizes to power genome-wide associations [3]. In order to gain tractability in the genetic causes of pediatric cancers for improved specific treatments, an alternative model that very closely recapitulates pediatric cancers at higher rates is needed.

The Pet Dog as a Genetic Model

The domesticated dog (*Canis lupus familiaris*) is an excellent model of human cancers for several reasons, including their easy accessibility and prominent status in diverse cultures [30].

For instance, over $40 billion (USD) is spent annually on dog health care [31] which is second only to humans in the level of health care received [32]. While laboratory animal models have long been used to study many diseases, recently, several studies have claimed that domestic dogs are the better animal model for human diseases [33]. This is partly due to dogs sharing over ~650 Mb of ancestral sequence with humans that is absent in mice, and them having DNA and protein sequence that is more similar to human than it is between human and mouse [26, 27]. Additionally, dog and human tumors demonstrate similar histopathology suggesting dogs and humans having complementary genetic features [33]. However, the greatest advantage of dog models is the result of their evolutionary history that involved at least two severe population bottlenecks [30]. Approximately 200 years ago most dog breeds were created by the selection of morphological and behavioral traits [34-36]. This was vastly accelerated during the controlled breeding practices of the Victorian era (circa 1830–1900), when crosses between breeds from divergent genetic lineages become highly desirable [18, 37]. Today's breeds are essentially isolated genetic populations whose genetic similarities and differences can be exploited to identify disease mutations [35].

Advantageously for any animal model of human diseases, dog cancer rates are very high, with NHL accounting for at least 10% of the over 6 million cancer diagnoses made each year [83% of all haematopoietic malignancies; 20, 38, 39]. Specifically, in DLBCL the incidence is estimated at 20 – 125 /100,000, suggesting national DLBCL rates of ~17,000 – 112,000 dogs [40-42]. The incidence of NHL is increasing in both dogs and humans and varies depending on age, histology, gender, and race [43]. In both humans and canines, activating pathway processes in lymphoproliferative diseases are shared, and evidence suggests that there are similarities in tumor microenvironment, clinical, cytological, and immunophenotypic between both species [44, 45]. In canines, as in humans, DLBCL is the leading identified histotype (44.4%) and is of primary B-cell

origin [31]. A seminal work in this field identified two primary expression patterns in DLBCL – this was observed in both humans and canines [32, 44, 46, 47]. And while the two differential expression patterns were not 100% concordant in terms of all genes within each group, canine and human lymphomas clustered *according to subtype, not species* [38, 47]. More recently, dogs are being used in clinical trials to study NF-kB-targeted therapeutics for human lymphoma due to the genetic similarity of the tumors [38]. Collectively, these studies identify molecular similarities in human DLBCL that suggest pet dogs are a highly representative model of DLBCL [47, 48].

Purpose

1. Identify differential gene expression in high-grade canine lymphoma tumor compared to locally adjacent normal tissue with cancer survival.
2. Identify differential DNA methylation in high-grade canine lymphoma tumors compared to locally adjacent normal tissue sample.

Conceptual Framework

We used the genetic single gene-two hit model to guide this work [17]. This framework combines rare and common variant theories with the addition of mutator/anti-mutator modulation to lead to disease development.

METHODS

Study Design

This study was designed to be a cross-sectional gene expression and DNA methylation analysis in pet dogs that presented to The Ohio State University Veterinary Medical Center and were diagnosed with High-grade B-cell lymphoma (HBL).

Sample Collection & Diagnosis

Animal care and sample collection were carried out in accordance with all applicable institutional, local, and national guidelines; dogs were under the care of licensed veterinarians, and participation did not influence decisions of care. Sample collection protocols were approved and reviewed by the Institutional Review Board of The Ohio State University and the Institutional Animal Care and Use Committee (IACUC) of The Ohio State University. Following an IACUC approved protocol, whole blood and tumor samples were collected by a licensed veterinarian or veterinary technician from pet dogs brought to The OSU Veterinary Medical Center for clinical evaluation. During routine clinical diagnosis, a 7-mL purple-top blood tube was obtained as well as tissue following tumor resection for pathology. Tumors were classified according to the modified WHO criteria based on morphology and immunophenotyped [49]. Based on the recent update of the 4th edition of the World Health Organization Classification of Haematopoietic and Lymphatic Tissues, the newly proposed classification of HBL represent more clinically aggressive cancer types and are associated with poor outcomes [50, 51]. Using this classification, we collected 71 high-quality biological samples from dogs diagnosed with three subtypes of canine HBL (cHBL): DLBCL (n = 52), Burkitt's-like lymphoma (BKL, n = 14), and Marginal Zone B-cell lymphoma (MZL, n = 5). Additionally, dogs diagnosed with Benign Follicular Hyperplasia (n = 7) after receiving lymph node resection were used as controls. Methodology used for pathology classification and diagnosis is standard in the field and has been used previously [52].

Biological Sample Preparation

DNA Isolation: Genomic DNA was isolated from leukocytes using 5 Prime ArchivePure DNA Blood Kits (Gaithersburg, MD) as per protocol. Samples were tested for quality using a 2% pre-stained ethidium bromide agarose gel with electrophoresis and subsequent fluorescence. Additional quality control included measurements with the Nanodrop 2000 and Qubit 3.0

Fluorometer readings per manufacturer's protocols. Results demonstrated visualized DNA as compact, high-molecular-weight bands with no low-molecular-weight smears and OD260/280 ratios ≤ 1.8 and OD260/230 ratios ≤ 2.0, thus indicating intact high-quality genomic DNA.

RNA Isolation: Total RNA was extracted from each tumor and locally adjacent normal tissue sample using TRIzol reagent (Invitrogen, San Diego, CA, USA). RNA was isolated from biopsy cells recovered from cryopreservation using the RNAeasy Mini Kit and QIAshredder (QIAGEN, Valencia, California) [as performed previously in 53]. RNA concentration was determined using NanoDrop ND-1000 UV-Vis spectrophotometer (NanoDrop Technologies, Wilmington, Delaware), and quality was measured using a 2100 Bioanalyzer (Agilent, Santa Clara, California). All the samples included in the gene expression profiling experiment were suitable for microarray analysis based on RNA quality (RIN > 7.0).

Array Hybridization & Sequencing

Affymetrix GeneChip Genome 2.0 Microarray: Using total RNA (isolation described above) from tumor and locally adjacent normal tissue samples, this microarray was processed at The Ohio State University Comprehensive Cancer Center Genomics Core (an approved Affymetrix laboratory site) for canine gene expression profiling. The array consists of 42,900 probe sets with 11 probe pairs per gene sequence. Oligonucleotide probe length for each probe is a 25-mer. For *C. familiaris*, there are 18,000 reference sequences, representing 20,000 non-redundant predicted genes. The sequence information provided in this array is from public data sources, including GenBank, UniGene, mRNAs, and the BROAD-1 annotation dataset. Collected samples were prepared following the standard Affymetrix protocol, as indicated by the manufacturer. Samples were hybridized to Affymetrix Canine 2.0 gene chips (Affymetrix, Santa Clara, California). Approximately 2.5 μg of RNA were labeled using the Affymetrix labeling protocol (Affymetrix,

Santa Clara, CA). The cRNA samples were then hybridized to Canine 2 gene expression chips as described [54]. The Canine 2.0 gene expression chip contains 43,000 annotated sequences derived from the 7.5× canine genome [55]. These represent virtually every known gene and a complement of expressed sequence tags that provide strong redundancy for expression profiling.

MethylCap-seq Library Generation and Sequencing: For DNA methylation analysis, we selected MethylCap-seq. This is a cost-efficient, genome-wide, highly reproducible, high-throughput, less cumbersome method than other traditional techniques used for interrogating methylated regions [56-58]. After fragmentation, methylated DNA is captured with the high affinity methyl-CpG binding domain of human MBD2 protein and eluted in a step-wise manner indicative of methyl-CpG density [57, 58]. Consequent analysis is performed on the enriched fragments by massively parallel sequencing [56, 57]. MethylCap-Seq was performed at The Ohio State University Comprehensive Cancer Center's Genomics Shared Resource. Briefly, DNA was quantified by Qubit fluorometric quantitation (Life Technologies, Grand Island, NY). 1-1.3 μg DNA was subjected to fragmentation using a Covaris S2 Adaptive Acoustic instrument (Woburn, MA) to average fragment size of 100 – 250 bp. Methylated DNA enrichment fragments were enriched using the Diagenode AutoMethylCap Kit customized for the Diagenode SX-8G IP-Star Compact Automated System (Denville, NJ). 1 ng of the resultant enriched and ethanol precipitated methylated DNA was used to generate an Illumina-compatible sequencing library using the Kapa Hyper Prep Kit (Fremont, CA). Library fragments (average fragment size is 300-400 nt, including adapters) were amplified using Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB) for 8 PCR cycles. Library material was purified with Agencourt AMPure XP beads (1:1 volume; Beckman Coulter, Inc., Indianapolis, IN). Sequencing was performed on the Illumina HiSeq 2500

(San Diego, CA) using the single-end 50-bp approach (Version 3 chemistry) to a depth of at least 40 million passed filtered reads.

MethylCap-seq Read Alignment & Assignment of Methylation Values from Reads:

Sequencing reads were received in fastq format. Duplicates (i.e. all reads with the same sequence data) among passed filter (using default Illumina settings) sequencing reads were collapsed. Collapsed reads were quality trimmed using a quality score cutoff of 33 using the fastx toolkit. Reads were then aligned to the canine genome CanFam3 using bowtie [version 0.12.7, 43]. DNA fragment profiles were extracted for every sample from Bioanalyzer data using a custom python script. These fragment profiles and the aligned reads were employed to assign methylation values to every CpG in the human genome using PrEMeR-CG [59]. Finally, differential methylation was called between Day 1 and Day 3 samples using paired MethMAGE on the first exons of protein coding genes [60]. We focused on the methylation of first exons of annotated protein coding genes, which has been shown to be especially relevant for regulation of gene expression [61]. Given that this is the first genome-wide analysis of canine DNA methylation changes, we applied stringent thresholds to the data. We removed any regions in which MethMAGE failed (due to absence of methylation data), required that the maximum of the methylation signal in the two groups within a region was above its median across all regions, and applied Benjamini Hochberg FDR [BHFDR, 46] with a q-value cutoff of 0.01.

Statistical & Laboratory Analysis

Software: The raw probe level signal intensity data from the Affymetrix array was analyzed using JMP Genomics [54]. We performed the Basic Expression Workflow that included background adjustment, quantile normalization, and probe-set summarization with Median Polish

analysis for pre-processing of the raw data [as recommended by JMP, 55]. Pre-processing was divided into two phases: Filtering and Quality Control (QC) per recommended protocol.

Subsequent to the filtering and QC, a final dataset was analyzed using ~40,000 probes. Using the finalized dataset, we conducted an Analysis of Variance (ANOVA) to determine statistically significant changes in the probe level signal intensities between genes associated with specific subtypes of B-cell lymphoma. We set the categorical variables for different lymphoma genes (for Hierarchical Clustering) and days of survival (for Kaplan Meier Analysis). We conducted a false discovery rate (FDR; allowing tolerance of a limited number of tests to have a Type I error) which was controlled at 0.001 using the Benjamini-Hochberg FDR to obtain p-values. Subsequent lists of differentially expressed genes were filtered based on fold change and overall p-value (additionally calling overlapping genes in cHBL, DLBCL, and BFH using mean expression with fold change $\leq$ -3 and $\geq$ 3). For those gene transcripts that were significantly different (unadjusted p-value $\leq$ 0.05), a post-hoc Tukey HSD test was performed to determine which groups differed significantly at p-value $\leq$ 0.05 [56]. As applied in our dataset, the Tukey HSD is conservative as the sample sizes differ [57].

Clustering: Hierarchical clustering was performed following ANOVA, using probe-sets filtered as noted above. The values of gene expression were median centered and normalized to a standard deviation of 1. Using the results, we generated a heat map to demonstrate sample relationships in terms of gene expression. K-means clustering was also performed following ANOVA to cluster the genes that have similar expression patterns. The two-level approach was applied with analysis of cHBL and BFH groups. K-means clustering is a partition clustering, and clusters the data into groups preset by the researcher. Using the set of observed survival times, we utilized the Kaplan-Meier survival curve, which shows the survival time in a “step function”.

Gene Ontology Analysis: Data was uploaded to QIAGEN's Ingenuity Pathway Analysis (IPA, QIAGEN Redwood City, http://www.qiagen.com/ingenuity) for biological annotation. All analyses were conducted with IPA defined defaults. The p-value was calculated using a right-tailed Fisher Exact Test that, for a given annotation process, is calculated by considering both the number of focus genes within that process and the total number of genes that are associated with that process in the reference set [62]. Overall, p-values ≤ 0.05 indicate a statistically significant, non-random association. P-values listed for IPA analysis are uncorrected except where noted. We used uncorrected p-values as starting points for further downstream analysis because Fisher's exact test can be overly conservative with small sample sizes as we have here [63, 64]. For specific hypothesis testing, we used the Benjamini-Hochberg method of correction [62, 65].

Methylation-specific PCR for Validation: To validate the DNA methylation sequencing results, we used a Methylation-specific PCR (MSP). MSP was designed based on publicly available databases [66-69]. The technique is summarized as follows: (1) Two restriction enzymes (RE) with cleavage sites overlapping a CpG island in the region of interest are selected. Importantly, one RE is methylation sensitive while the other (an isoschizomer) is not. (2) After digestion with the REs, a PCR using specific primer sets for both unmethylated and methylated DNA is performed. MSP is a rapid measure for assessing DNA methylation status in CpG island.

RESULTS

Gene Expression

Affymetrix Array Results: After performing data clean-up and normalization using JMP Genomics protocol [70], we identified all differentially expressed genes with fold-change>3 and

$p \leq 0.001$ in the cHBL (n = 71) and Benign Follicular Hyperplasia (n = 7) comparison. In total, we identified 1,180 differentially expressed transcripts, representing 464 genes.

Gene Symbol	P-value range; Sig. transcripts* (total)
CDH1 Cadherin 1	≤0.0001 - 0.3813; 5(7)
ITGA6 Integrin Subunit Alpha	≤0.0001 - 0.286; 3(4)
KCNJ8 Potassium Channel	≤0.0001 - 0.0006; 4(4)
LAMB1 Laminin Subunit Beta	≤0.0001 - 0.4032; 6(9)
PEG3 Paternally Expressed 3	≤0.0001 - 0.1793; 4(5)
RAMP2 Receptor Mod. Protein	≤0.0001 - 0.8863 ; 3(4)

Table 2: Genes Significantly Correlated with Survival in High-Grade Canine Lymphomas. The gene names and symbols we identified are presented with the range of p-values, the total number of transcript represented by the gene, and the number of significant transcripts.

Gene Expression Correlated with Survival in Tumor Compared to Locally Adjacent Normal Tissue Sample: For a subset of the pet dogs diagnosed with cHBL (n = 45) survival data was available. Using K-mean clustering [an unsupervised learning algorithm; 71], we separated the samples into two groups based on the combined expression of the above 464 genes (high and low expressers). We then correlated each expression-based sample group with survival time ($p \leq 0.05$) and generated a list of significant genes for each group. Only 6 genes overlapped in the lists from high and low expressers after selecting significant transcripts (fold-change > 3 and $p < 0.001$) in subset of dogs and correlating with survival time (Table 2).

Region	DMRs	Percent Significant	Percent HYPOmethylated	Percent HYPERmethylated
All Regions	1,429	63.51%	33.45%	13.37%
CPG Island	20,078	42.56%	65.00%	25.33%
Promoter	1,190	52.89%	65.38%	18.40%
First Exon	633	28.13%	67.93%	28.91%
All Exons	3,068	19.11%	35.56%	60.76%

Table 3: Differentially Methylated DNA regions in DLBCL & Locally Adjacent Normal Tissue. We report the overall DNA methylated regions in 5 different features specifically selected using MeDIP. Presented are the total number of significant and percentage of significant regions, percentage of hypomethylated regions and percentage of hypermethylated regions in DLBCL.

DNA Methylation

Differentially Methylated Regions *in Tumor Compared to Locally Adjacent Normal Tissue Sample*:

In total, we identified 4,891 differentially methylated genes representing 3 major elements: Promoter (DMRs = 1,190), First Exon (DMRs = 633), and All Exons (DMRs = 3,060; Table 3). We focused the remaining analysis on the first exon since DNA methylation of this element is most closely associated with gene transcription. Using only the 633 DMRs, we identified the top five canonical pathways that were overly represented by number of genes from that pathway in our significant DMRs. These included: G-protein coupled receptor signaling, cAMP-mediated signaling, hepatic fibrosis, glucocorticoid receptor signaling, and communication between innate and adaptive immune cells.

Ingenuity Pathway Analysis

Statistically significant canonical pathways were analyzed using the Ingenuity Pathway Analysis (IPA) for the differentially methylated regions of the first exon. After selecting the significant transcripts (fold change > 2 and $p < 0.001$) the top five most significant genes were determined in both DLBCL and locally adjacent normal tissues (LANT). In this comparison, the five significant genes in the LANT group signified as having lowest fold

DLBCL DNA Methylation Values					
Lowest Fold Change			**Highest Fold Change**		
Gene Symbol	**Fold Change**	**P-value**	**Gene Symbol**	**Fold Change**	**P-value**
MIR449B	11,743	7.25×10^{-4}	***ARX***	791,960	8.77×10^{-6}
OCLN	1,094	2.32×10^{-5}	***PPP1R2***	138,715	4.87×10^{-4}
MIR8880	205	1.97×10^{-13}	***HPGD***	127,013	1.32×10^{-4}
MIR506	199	4.91×10^{-6}	***MAOB***	66,934	2.78×10^{-5}
TNNT2	18	2.26×10^{-20}	***SRP68***	56,426	2.14×10^{-3}

Table 4: Top 5 Genes with the Largest and Smallest Fold-Change in DLBCL Compared to Normal Tissue.

change in DLBCL. Table 4 shows the top five significant genes for LANT and DLBCL comparison.

In addition, the top five canonical pathways in both LANTs and DLBCL were determined using the same significant transcripts (fold change > 2 and $p < 0.001$) for each group. In LANTs, the top five canonical pathways were glucocorticoid receptor signaling, neuroinflammation signaling pathway, hepatic fibrosis/hepatic stellate cell activation, cAMP-mediated signaling, and corticotropin releasing hormone signaling. In DLBCL, the top five canonical pathways included G-protein coupled receptor signaling, hepatic fibrosis/hepatic stellate cell activation, cAMP-mediated signaling, glutamate dependent acid resistance, and embryonic stem cell differentiation into cardiac lineages. Between the two groups, cAMP-mediated signaling and hepatic fibrosis/hepatic stellate cell activation were the pathways seen in both groups.

Comparison of Expression and DNA Methylation Results

We found that *CDH1* was more hypermethylated compared to the control at the first exon. Hypermethylation suggests decreased expression, and this was confirmed in our expression data. Through Treehouse, the publicly available pediatric expression dataset, expression values in DLBCL were analyzed. We found that decreased expression of *CDH1* was a predictor of death by the median survival. These findings were validated in 91 pediatric tumors with survival data. Additionally, decreased *CDH1*

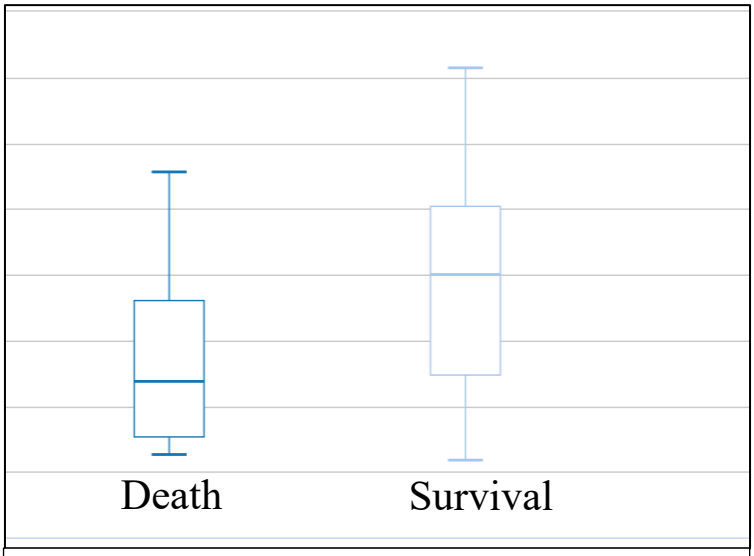

Figure 1: Expression Values for Pediatric DLBCL of *CDH1*. Those who expressed lower levels of CDH1 were more likely to die from their disease than higher expressers.

expression is associated with poor survival in multiple types of cancers. Overall, the expression of this gene was significantly associated with survival in pediatric lymphoma ($p = 0.013$; see Figure 1).

DISCUSSION

All of the 6 prognostic genes identified through differential gene expression, have biological roles that are significant for cancer development. *CDH1* is a known tumor suppressor gene, where knockdown or silencing of this gene can lead to increase in cancer proliferation and invasion [72]. *ITGA6* is known to promote tumorigenesis, causing alternative splicing in transcript variants [73]. *KCNJ8* regulate integral membrane proteins and influences potassium channel which are responsible for variety of physiological responses [74]. *LAMB1* controls cell adhesion, differentiation, migration, signaling, and cancer metastasis [75]. *PEG3* is associated with cell proliferation and apoptosis as well as potential for suppressing tumor growth [76]. Lastly, *RAMP2* is a gene that suppresses metastasis and maintain vascular integrity [77].

Utilizing IPA, the top five significant genes in LANTs and DLBCL of first exon were determined. In LANTs, the top five significant genes were *MIR449B*, *OCLN*, *MIR8880*, *MIR506*, and *TNNT2*. The *MIR* genes are microRNAs, which are involved in post-transcriptional regulation of gene expression by affecting the translation of mRNAs. MicroRNAs can recognize the target mRNAs and cause translational inhibition or destabilization of that mRNA. The *OCLN* gene encodes an integral membrane protein, affecting platelet function tests and is highly expressed in the thyroid gland. *TNNT2* encodes the protein in tropomyosin-binding subunit, which is highly expressed in left ventricle of the heart. It influences the regulation of cholesterol (HDL), neurofibrillary tangles, and troponin T.

In addition, the most common canonical pathways seen in this group include glucocorticoid receptor signaling, neuroinflammation signaling pathway, hepatic fibrosis/hepatic stellate cell

activation, cAMP-mediated signaling, and corticotropin releasing hormone signaling. Glucocorticoid receptor signaling significantly affects anti-inflammatory, immunosuppressive, and inflammatory-disease modulation. Glucocorticoids, like other corticosteroids, have been used for treatment of cancers for their anti-proliferative and antiangiogenic properties [78]. As with glucocorticoids, neuroinflammation signaling pathway is involved with inflammation. It is known that cancer inflammation can promote or exacerbate cancer development [79]. Hepatic stellate cell activation is also a pro-inflammatory state, where prolonged hepatic injury can cause the accumulation of activated hepatic stellate cells. This can lead to irreversible state of liver cirrhosis. Hepatic stellate cell clearance has also been associated with apoptosis [80]. The c-AMP-mediated signaling regulates cell growth, cell differentiation, gene transcription, and protein expression. As a second messenger pathway, it is associated with multiple diseases such as inflammation, cancer, and myocardial atrophy [81]. Lastly, corticotropin releasing hormone signaling pathway affects the hypothalamic-pituitary-adrenal (HPA) axis and is a principle pathway for stress response. Corticotropin releasing hormone is also known to employ the cAMP cascade, which has an effect on cancer development [82]. Some of the diseases and disorders that were most commonly affected by the LANTs group were metabolic disease, endocrine system disorders, GI disease, organismal injury and abnormalities, and connective tissue disorders.

Within the DLBCL group, the top five significant genes were *ARX*, *PPP1R2*, *HPGD*, *MAOB*, and *SRP68*. *ARX* is a highly expressed gene in the ovary and is a lipoprotein affecting gene that is expressed during development. *PPP1R2* is one of the main eukaryotic serine/threonine phosphates that is highly expressed in the testis. *HPGD* encodes enzyme that is responsible for metabolism of prostaglandins and is highly expressed in the bladder. It influences the platelet function tests, venous thromboembolism, insulin like growth factor I, blood pressure, and body fat

distribution. *MAOB* encodes the protein that belongs to flavin monoamine oxidase family. It is highly expressed in the ovary and affects aorta, smoking, and bipolar disorder. Lastly, *SRP68* is a gene that encodes a subunit of signal recognition particle (*SRP*). It is highly expressed in cells especially the transformed fibroblasts.

The top canonical pathways seen in this set of genes include the G-protein coupled receptor signaling, hepatic fibrosis/hepatic stellate cell activation, cAMP-mediated signaling, glutamate dependent acid resistance, and embryonic stem cell differentiation into cardiac lineages. G-protein coupled receptor signaling regulate the cell cycle progression. Recent evidence shows that G-protein coupled receptors have effect on cancer development, including vascular remolding, invasion, and migration [83]. The glutamate dependent acid resistance is utilized by neutralophilic bacteria to combat environmental situations with low pH. These bacteria that affect humans, utilize this pathway to cope with acid stress, or to protect cells from acidic environment such as the intestines [84]. Lastly, the embryonic stem cells have the potential to be differentiated into cardiac progenitor cells, to utilize as a cell-based therapy for heart disease [85]. The most common diseases associated with the DLBCL group include cancer, organismal injury and abnormalities, neurological disease, skeletal and muscular disorders, and hereditary disorders. Between the LANTs and DLBCL, the only overlapping disorder was organismal injury and abnormalities.

CONCLUSION

This study is one of the first to provide an alternative model for aggressive lymphomas in children. Pet dog models are valuable for higher understanding of the genetic basis of pediatric high-grade lymphomas. Through studying of differential gene expression and differential DNA methylation, *CDH1* was identified as the overlapping gene. In both canine and pediatric high-grade lymphomas, *CDH1* is indicative of decreased survival. More research to understand the genetic

mechanism of *CDH1* is still necessary to fully understand its biomarker relevance for future clinical application.

AWKNOWLEDGEMENT

First and foremost, I would like to thank The Ohio State University College of Nursing for the funding and the opportunity to participate in this project, as well as all of the dogs and dog owners who agreed to be a part of this research. I am especially indebted to my advisor, Dr. Jennie Rowell, who supported me throughout the research process, and have worked actively to help me reach my academic and career goals. I would also like to show my gratitude to Elizabeth McNamara and Dr. Shannon Gillespie for their contributions to this thesis. Finally, I would like to thank my parents and my fiancé, Evan, for their unending support.

REFERENCES

1. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer Statistics, 2017.* CA Cancer J Clin, 2017. **67**(1): p. 7-30.
2. Hudson, M.M., et al., *Lessons from the past: opportunities to improve childhood cancer survivor care through outcomes investigations of historical therapeutic approaches for pediatric hematological malignancies.* Pediatr Blood Cancer, 2012. **58**(3): p. 334-43.
3. Capitini, C.M., et al., *Immunotherapy in pediatric malignancies: current status and future perspectives.* Future Oncol, 2014. **10**(9): p. 1659-78.
4. Hudson, M.M., et al., *Health status of adult long-term survivors of childhood cancer: a report from the Childhood Cancer Survivor Study.* JAMA, 2003. **290**(12): p. 1583-92.
5. Maurice-Stam, H., et al., *Longitudinal assessment of health-related quality of life in preschool children with non-CNS cancer after the end of successful treatment.* Pediatr Blood Cancer, 2008. **50**(5): p. 1047-51.
6. Sankila, R., et al., *Geographical comparison of cancer survival in European children (1988-1997): report from the Automated Childhood Cancer Information System project.* Eur J Cancer, 2006. **42**(13): p. 1972-80.
7. Voûte, P.A., et al., *Cancer in children: clinical management*. 2012: Springer Science & Business Media.
8. Blaauwbroek, R., et al., *Late effects in adult survivors of childhood cancer: the need for life-long follow-up.* Ann Oncol, 2007. **18**(11): p. 1898-902.
9. http://www.cancer.gov/types/childhood-cancers/child-adolescent-cancers-fact-sheet.
10. Stam, H., et al., *Health-related quality of life in children and emotional reactions of parents following completion of cancer treatment.* Pediatr Blood Cancer, 2006. **47**(3): p. 312-9.
11. Group, C.s.O. *Long-Term Follow-Up Guidelines for Survivors of Childhood, Adolescent, and Young Adult Cancers* 2018; Available from: http://www.survivorshipguidelines.org/.
12. http://www.who.int/mediacentre/factsheets/fs178/en/.
13. http://seer.cancer.gov/faststats/selections.php?#Output.
14. Esteller, M., *Aberrant DNA methylation as a cancer-inducing mechanism.* Annu Rev Pharmacol Toxicol, 2005. **45**: p. 629-56.
15. Olsson, M., et al., *Genome-wide methylation profiling identifies novel methylated genes in neuroblastoma tumors.* Epigenetics, 2016. **11**(1): p. 74-84.
16. Mai, H., et al., *Hypermethylation of p15 gene associated with an inferior poor long-term outcome in childhood acute lymphoblastic leukemia.* J Cancer Res Clin Oncol, 2016. **142**(2): p. 497-504.
17. Hussain, S., *A new conceptual framework for investigating complex genetic disease.* Front Genet, 2015. **6**: p. 327.
18. Boyko, A.R., et al., *A simple genetic architecture underlies morphological variation in dogs.* PLoS Biol, 2010. **8**(8): p. e1000451.
19. Percy CL, S.M., Linet M, et al.: Lymphomas and reticuloendothelial neoplasms. In: Ries LA, Smith MA, Gurney JG, et al., eds.: Cancer incidence and survival among children and adolescents: United States SEER Program 1975-1995. Bethesda, Md: National Cancer Institute, SEER Program, 1999. NIH Pub.No. 99-4649., pp 35-50. Also available online. Last accessed April 03, 2014.
20. Sandlund, J.T., J.R. Downing, and W.M. Crist, *Non-Hodgkin's lymphoma in childhood.* N Engl J Med, 1996. **334**(19): p. 1238-48.
21. http://www.cancer.gov/cancertopics/types/hodgkin.

22. Chiu, B.C. and N. Hou, *Epidemiology and etiology of non-hodgkin lymphoma.* Cancer Treat Res, 2015. **165**: p. 1-25.
23. Swerdlow, S.H., International Agency for Research on Cancer., and World Health Organization., *WHO classification of tumours of haematopoietic and lymphoid tissues*. 4th ed. World Health Organization classification of tumours. 2008, Lyon, France: International Agency for Research on Cancer. 439 p.
24. *SEER Cancer Stat Facts: Non-Hodgkin Lymphoma*. [cited 2017 May, 22]; Available from: http://seer.cancer.gov/statfacts/html/nhl.html
25. *Cancer Facts & Figures 2018*. 2018; Available from: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf.
26. Richards, K.L., et al., *Gene profiling of canine B-cell lymphoma reveals germinal center and postgerminal center subtypes with different survival times, modeling human DLBCL.* Cancer research, 2013. **73**(16): p. 5029-5039.
27. Miles, R.R., et al., *Pediatric diffuse large B-cell lymphoma demonstrates a high proliferation index, frequent c-Myc protein expression, and a high incidence of germinal center subtype: Report of the French-American-British (FAB) international study group.* Pediatric Blood & Cancer, 2008. **51**(3): p. 369-374 6p.
28. Sandlund, J.T. and M.G. Martin, *Non-Hodgkin lymphoma across the pediatric and adolescent and young adult age spectrum.* Hematology Am Soc Hematol Educ Program, 2016. **2016**(1): p. 589-597.
29. https://www.cancer.gov/types/childhood-cancers/child-adolescent-cancers-fact-sheet. Available from: https://www.cancer.gov/types/childhood-cancers/child-adolescent-cancers-fact-sheet.
30. Rowell, J.L., D.O. McCarthy, and C.E. Alvarez, *Dog models of naturally occurring cancer.* Trends Mol Med, 2011. **17**(7): p. 380-8.
31. Ferraresso, S., et al., *Epigenetic silencing of TFPI-2 in canine diffuse large B-cell lymphoma.* PLoS One, 2014. **9**(4): p. e92707.
32. Alizadeh, A.A., et al., *Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.* Nature, 2000. **403**(6769): p. 503-11.
33. Bushell, K.R., et al., *Genetic inactivation of TRAF3 in canine and human B-cell lymphoma.* Blood, 2015. **125**(6): p. 999-1005.
34. Lango Allen, H., et al., *Hundreds of variants clustered in genomic loci and biological pathways affect human height.* Nature, 2010. **467**(7317): p. 832-8.
35. Lian, C.G., et al., *Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma.* Cell, 2012. **150**(6): p. 1135-46.
36. Ngollo, M., et al., *Epigenetic modifications in prostate cancer.* Epigenomics, 2014. **6**(4): p. 415-26.
37. Rimbault, M., et al., *Derived variants at six genes explain nearly half of size reduction in dog breeds.* Genome Res, 2013. **23**(12): p. 1985-95.
38. Mudaliar, M.A., et al., *Comparative gene expression profiling identifies common molecular signatures of NF-kappaB activation in canine and human diffuse large B cell lymphoma (DLBCL).* PLoS One, 2013. **8**(9): p. e72591.
39. Foundation, A.C. 2018; Available from: http://www.acfoundation.org/faqs/.
40. Richards, K.L. and S.E. Suter, *Man's best friend: what can pet dogs teach us about non-Hodgkin's lymphoma?* Immunol Rev, 2015. **263**(1): p. 173-91.

41. Dobson, J.M., et al., *Canine neoplasia in the UK: estimates of incidence rates from a population of insured dogs.* J Small Anim Pract, 2002. **43**(6): p. 240-6.
42. American Veterinary Medical, A., *U.S. pet ownership & demographics sourcebook*. 2012.
43. Ito, D., A.M. Frantz, and J.F. Modiano, *Canine lymphoma as a comparative model for human non-Hodgkin lymphoma: recent progress and applications.* Vet Immunol Immunopathol, 2014. **159**(3-4): p. 192-201.
44. Aresu, L., et al., *Canine indolent and aggressive lymphoma: clinical spectrum with histologic correlation.* Vet Comp Oncol, 2013.
45. Sokolowska, J., et al., *Proliferation activity in canine lymphomas.* Pol J Vet Sci, 2012. **15**(4): p. 727-34.
46. Lenz, G., et al., *Stromal gene signatures in large-B-cell lymphomas.* N Engl J Med, 2008. **359**(22): p. 2313-23.
47. Richards, K.L., et al., *Gene profiling of canine B-cell lymphoma reveals germinal center and postgerminal center subtypes with different survival times, modeling human DLBCL.* Cancer Res, 2013. **73**(16): p. 5029-39.
48. Sato, M., et al., *Evaluation of the prognostic significance of BCL6 gene expression in canine high-grade B-cell lymphoma.* Vet J, 2012. **191**(1): p. 108-14.
49. Asou, N., *Myeloid neoplasms in the World Health Organization 2016 classification.* Rinsho Ketsueki, 2017. **58**(10): p. 2178-2187.
50. Ott, G., *Aggressive B-cell lymphomas in the update of the 4th edition of the World Health Organization classification of haematopoietic and lymphatic tissues: refinements of the classification, new entities and genetic findings.* Br J Haematol, 2017. **178**(6): p. 871-887.
51. Jiang, M., N.N. Bennani, and A.L. Feldman, *Lymphoma classification update: B-cell non-Hodgkin lymphomas.* Expert Rev Hematol, 2017. **10**(5): p. 405-415.
52. Briggs, J., et al., *A compendium of canine normal tissue gene expression.* PLoS One, 2011. **6**(5): p. e17107.
53. Li, X., et al., *Evaluation of microRNA Expression in Patients with Herpes Zoster.* Viruses, 2016. **8**(12).
54. Burton-Wurster, N., et al., *Genes in canine articular cartilage that respond to mechanical injury: gene expression studies with Affymetrix canine GeneChip.* J Hered, 2005. **96**(7): p. 821-8.
55. Lindblad-Toh, K., et al., *Genome sequence, comparative analysis and haplotype structure of the domestic dog.* Nature, 2005. **438**(7069): p. 803-19.
56. Rodriguez, B.A., et al., *Methods for high-throughput MethylCap-Seq data analysis.* BMC Genomics, 2012. **13 Suppl 6**: p. S14.
57. De Meyer, T., et al., *Genome-wide DNA methylation detection by MethylCap-seq and Infinium HumanMethylation450 BeadChips: an independent large-scale comparison.* Sci Rep, 2015. **5**: p. 15375.
58. Brinkman, A.B., et al., *Whole-genome DNA methylation profiling using MethylCap-seq.* Methods, 2010. **52**(3): p. 232-6.
59. Frankhouser, D.E., et al., *PrEMeR-CG: inferring nucleotide level DNA methylation values from MethylCap-seq data.* Bioinformatics, 2014. **30**(24): p. 3567-74.
60. Ayyala, D.N., et al., *Statistical Methods for Detecting Differentially Methylated Regions Based on MethylCap-Seq Data.* Brief Bioinform, 2015.
61. Brenet, F., et al., *DNA methylation of the first exon is tightly linked to transcriptional silencing.* PLoS One, 2011. **6**(1): p. e14524.

62. https://www.ingenuity.com/wp-content/themes/ingenuity-qiagen/pdf/ipa/functions-pathways-pval-whitepaper.pdf.
63. Liddell, D., *Practical Tests of 2 x 2 Contingency Tables.* Journal of the Royal Statistical Society. Series D (The Statistician), 1976. **25**(4): p. 295-304.
64. D'Agostino, R.B., W. Chase, and A. Belanger, *The Appropriateness of Some Common Procedures for Testing the Equality of Two Independent Binomial Populations.* The American Statistician, 1988. **42**(3): p. 198-202.
65. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society. Series B (Methodological), 1995. **57**(1): p. 289-300.
66. http://www.restrictionmapper.org/. Available from: http://www.restrictionmapper.org/.
67. http://rebase.neb.com/cgi-bin/msrecget. Available from: http://rebase.neb.com/cgi-bin/msrecget.
68. http://biotools.nubic.northwestern.edu/OligoCalc.html. Available from: http://biotools.nubic.northwestern.edu/OligoCalc.html
69. http://www.urogene.org/cgi-bin/methprimer/methprimer_results.cgi. Available from: http://www.urogene.org/cgi-bin/methprimer/methprimer_results.cgi
70. https://www.jmp.com/en_ch/life-sciences/step-by-step-guides/basic-expression-analysis.html. Available from: https://www.jmp.com/en_ch/life-sciences/step-by-step-guides/basic-expression-analysis.html.
71. Fong, S., et al., *Towards enhancement of performance of K-means clustering using nature-inspired optimization algorithms.* ScientificWorldJournal, 2014. **2014**: p. 564829.
72. *CDH1 cadherin 1 [Homo sapiens (human)]*. 2018, National Library of Medicine (US), National Center for Biotechnology Information: Bethesda, MD.
73. *ITGA6 integrin subunit alpha 6 [Homo sapiens (human)]*. 2018, National Library of Medicine (US), National Center for Biotechnology Information: Bethesda, MD.
74. *KCNJ8 potassium voltage-gated channel subfamily J member 8 [Homo sapiens (human)]*. 2018, National Library of Medicine (US), National Center for Biotechnology Information: Bethesda, MD.
75. *LAMB1 laminin subunit beta 1 [Homo sapiens (human)]*. 2018, National Library of Medicine (US), National Center for Biotechnology Information: Bethesda, MD.
76. *PEG3 paternally expressed 3 [Homo sapiens (human)]*. 2018, National Library of Medicine (US), National Center for Biotechnology Information: Bethesda, MD.
77. *RAMP2 receptor activity modifying protein 2 [Homo sapiens (human)]*. 2018, National Library of Medicine (US), National Center for Biotechnology Information: Bethesda, MD.
78. Kadmiel, M. and J.A. Cidlowski, *Glucocorticoid receptor signaling in health and disease.* Trends Pharmacol Sci, 2013. **34**(9): p. 518-30.
79. Nakagawa, H. and S. Maeda, *Inflammation- and stress-related signaling pathways in hepatocarcinogenesis.* World J Gastroenterol, 2012. **18**(31): p. 4071-81.
80. Tsuchida, T. and S.L. Friedman, *Mechanisms of hepatic stellate cell activation.* Nat Rev Gastroenterol Hepatol, 2017. **14**(7): p. 397-411.
81. Yan, K., et al., *The cyclic AMP signaling pathway: Exploring targets for successful drug discovery (Review).* Mol Med Rep, 2016. **13**(5): p. 3715-23.
82. Grammatopoulos, D.K., *Insights into mechanisms of corticotropin-releasing hormone receptor signal transduction.* Br J Pharmacol, 2012. **166**(1): p. 85-97.

83. Yu, S., et al., *The Role of G Protein-coupled Receptor Kinases in Cancer.* Int J Biol Sci, 2018. **14**(2): p. 189-203.
84. Lund, P., A. Tramonti, and D. De Biase, *Coping with low pH: molecular strategies in neutralophilic bacteria.* FEMS Microbiol Rev, 2014. **38**(6): p. 1091-125.
85. Lei, I.L., L. Bu, and Z. Wang, *Derivation of cardiac progenitor cells from embryonic stem cells.* J Vis Exp, 2015(95): p. 52047.